



b-jet identification algorithms and performance in the ATLAS experiment

Kazuya Mochizuki, on behalf of the ATLAS collaboration

CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France

kazuya.mochizuki@cern.ch

Abstract

For many physics analyses at the LHC, the ability to identify jets containing *B*-hadrons (*b*-tagging) is one of the most crucial ingredients. After a brief review of the *b*-tagging algorithms developed in the ATLAS experiment, we present the refined *b*-jet tagging efficiency measurement with a combinatorial likelihood approach, a novel *c*-jet tagging efficiency measurement using *W* boson production in association with a *c*-quark, and the updates of *c*- and light-jet tagging efficiency measurements with increased statistics.

Keywords: LHC, ATLAS, *b*-tagging

1. Introduction

Identification of jets originating from *b*-quark fragmentation (*b*-tagging) plays an important role in physics programs at the LHC. In particular, it is essential for many physics analyses such as property measurements of the top quark, searches for the Higgs boson decaying to a *b*-quark pair ($H \rightarrow b\bar{b}$), and searches for new physics phenomena which are expected to have *b*-quark(s) in their final state.

In this proceedings article, we report the current status of *b*-tagging performance in the ATLAS experiment [1] at the LHC. The tagging efficiencies of *b*-, *c*-, and light-jets¹ are measured using data collected by the ATLAS experiment during the LHC Run 1. In the following sections, the term “7 (8) TeV dataset” is used to refer to the dataset corresponding to an integrated luminosity of 4.6 (20.3) fb⁻¹ collected in the year 2011 (2012) at a center-of-mass energy of 7 (8) TeV.

The *b*-tagging algorithms are reviewed in section 2. Novel methods to derive the tagging efficiency of *b*- and *c*-jets from data, a process referred to as calibration, are presented in sections 3 and 4. Updated calibrations using

the 8 TeV dataset of *c*- and light-jets are reported in sections 5 and 6.

2. *b*-tagging algorithms in the ATLAS experiment

Three basic *b*-tagging algorithms are employed by the ATLAS experiment: impact parameter (IP) based tagging algorithms, secondary vertex (SV) based algorithms, and the decay chain reconstruction algorithm. The MV1 tagger combines these taggers' information using a neural network. MV1c has also been developed as a variant of MV1, reducing *c*-jet contamination in tagged jets with a modest increase of light-jet contamination. The efficiencies of *b*-, *c*-, and light-jets for MV1 70% *b*-jet tagging efficiency working point (WP) are shown in Fig 1. The *b*-tagging WPs are defined by the average *b*-jet tagging efficiency measured in the $t\bar{t}$ simulated events. Differences between data and simulation performance are corrected using scale factors (SF) derived in several dedicated calibration methods for *b*-, *c*-, and light-jets, which are discussed in the following sections 3–6.

3. Combinatorial likelihood *b*-jet calibration

In this section, a new method to measure *b*-jet tagging efficiency using a combinatorial likelihood approach with dileptonic $t\bar{t}$ production events in the 8 TeV dataset

¹Light-jet is defined as a jet corresponding to neither *b*-, *c*-, nor τ -jets.

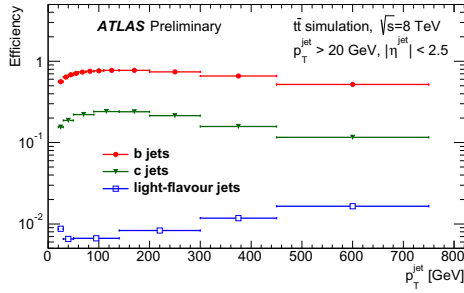


Figure 1: Efficiency of the MV1 tagger to select b -, c -, and light flavor jets as a function of p_T^{jet} for MV1 70% b -jet efficiency WP with selection of $p_T^{\text{jet}} > 20$ GeV and $|\eta^{\text{jet}}| < 2.5$, using $t\bar{t}$ simulated events [2].

is presented. In order to extract the efficiency corresponding to a certain b -tagging requirement, the following equations are used, for the case with exactly two jets in the event:

$$f_{2\text{tags}} = f_{bb}\epsilon_b^2 + f_{bj}\epsilon_j\epsilon_b + (1 - f_{bb} - f_{bj})\epsilon_j^2 \quad (1)$$

$$f_{1\text{tag}} = 2f_{bb}\epsilon_b(1 - \epsilon_b) + f_{bj}[\epsilon_j(1 - \epsilon_b) + (1 - \epsilon_j)\epsilon_b] + (1 - f_{bb} - f_{bj})2\epsilon_j(1 - \epsilon_j) \quad (2)$$

where ϵ_b (ϵ_j) is the b -jet (non b -jet) efficiency, $f_{1\text{tag}}$ ($f_{2\text{tags}}$) is the fraction of events with 1 (2) tagged jets, and f_{bb} (f_{bj}) is the fraction of events with a true bb (bj) jet pair. The b -tagging efficiency can be obtained by maximizing a likelihood function, taking both $f_{1\text{tag}}$ and $f_{2\text{tags}}$ from data, with f_{bb} , f_{bj} , and ϵ_j from simulation. Figure 2 shows the result of efficiency measurement in

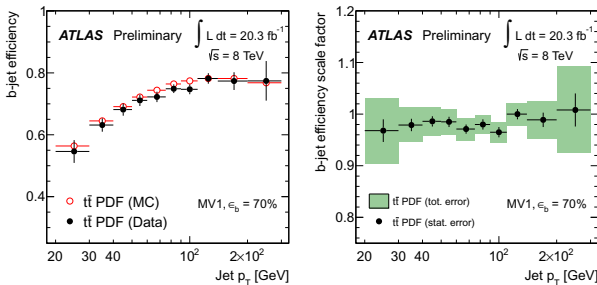


Figure 2: The p_T dependence of b -jet efficiencies (left) and b -jet efficiency SFs (right) for the MV1 algorithm at 70% b -jet efficiency WP. The error bars on data points in the left plot represents the total statistical and systematic uncertainties. In the right plot, the statistical only (black lines) and total errors (green shaded region) are shown [3].

both data and simulation and data-to-simulation SFs in ten p_T^{jet} bins. The SF varies in the range 0.965–1.008 with total uncertainties of 0.018–0.084.

4. $W + c$ calibration

This section describes a novel method to measure the tagging efficiencies of c -jets in the 7 TeV dataset

using $W + c$ production where selecting electrons from W decay ($W \rightarrow e\nu$) and c -jets with muon semi-leptonic decay ($c \rightarrow \mu X$) [4]. Since the soft muon carries a charge with the same sign as the c -quark, requiring $Q_\mu * Q_e = -1$ selects $W + c$ events with very high efficiency, where Q_μ and Q_e are the electric charges of the W 's electron and the charm's muon, respectively. Most of the background processes are evenly populated with events where the charges of the decay leptons have opposite sign (OS) or same sign (SS). Therefore, the number of c -jets is obtained by extracting the number of $W + c$ signal events as the difference between the number of events with opposite and with same charges (OS–SS). The c -jet efficiency, $\epsilon_{c(\mu)}^{\text{data}}$, is measured by calculating $\epsilon_{c(\mu)}^{\text{data}} = N_{Wc}^{b\text{-tagged}} / N_{Wc}$, where N_{Wc} and $N_{Wc}^{b\text{-tagged}}$ are the number of $W + c$ events before and after the b -tagging requirement. In Fig 3, the measured efficiency is shown for the MV1 tagger at 85%, 75%, 70%, and 60% b -jet tagging efficiency WPs. The c -jet efficiencies vary between 50% and 13% with total uncertainties of 3–10%, increasing with the tightness of the WP (lower b -jet efficiencies). The data-to-simulation SFs are found to be 0.75–0.92, with a 5–13% uncertainty.

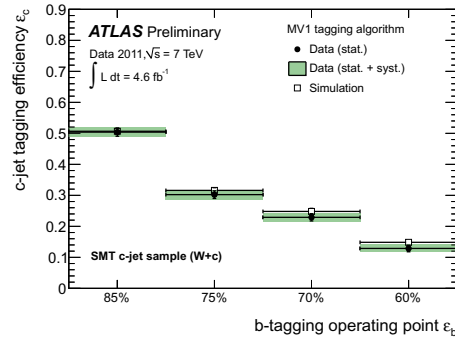


Figure 3: Comparison of the c -jet tagging efficiencies in data and the simulation of soft-muon-tagged c -jets derived for the MV1 algorithm using $W + c$ events [4].

5. D^* meson calibration

The c -jet efficiency is also measured using jets containing D^* ($D^{*\pm}$) mesons in the 8 TeV dataset, by comparing the yield of D^* before and after the b -tagging requirement. The $D^{*\pm} \rightarrow D^0\pi^\pm$ ($D^0 \rightarrow K^\mp\pi^\pm$) decay mode offers distinctive kinematic features, resulting in a modest combinatorial background. The contamination of D^* mesons originating from b -hadron decays is measured with a fit to the D^0 pseudo-proper time distribution [2]. Figure 4 shows the measured tagging efficiency of c -jets for data and simulation for MV1 at the 70% b -tagging efficiency WP. The data-to-simulation SF varies in the range 0.86–0.97 depending on p_T^{jet} , with a 8–15% uncertainty.

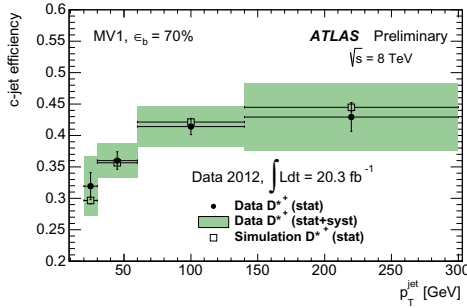


Figure 4: The c -jet tagging efficiency in data and simulation for jets containing D^* , for the 70% WP of the MV1 tagger [2].

6. Light-jet mistag rate calibration

The mistag rate is measured in the 8 TeV dataset. The mistag rate, defined as the fraction of light-jets which are tagged by the MV1 algorithm, is measured in an inclusive jet sample using the *negative tag* method [2]. Light-flavor jets are tagged as b -jets mainly because of the finite resolution of the inner detector and the presence of tracks from displaced vertices of long-lived particles or material interactions. Fake SVs and IPs of tracks in light jets are expected to distribute approximately symmetrically around the primary vertex while those from true SVs are not. The mistag rate is estimated by using the tracks' IP and/or SV decay lengths which are negatively signed. A correction has been applied for jet-direction mis-measurement for heavy flavor jets. For light-jets, another correction has been applied for tracks or SVs originating from long-lived particles and material interactions. The measured mistag rates in data and simulation for the MV1 tagging algorithm at 70% efficiency WP for central jets ($|\eta^{\text{jet}}| < 1.2$) are shown in Fig 5. For the chosen WP, the mistag rate ranges from 0.5% to 2.5%. The data/simulation SFs are slightly larger than unity, with relative total uncertainties ranging 15%–43%.

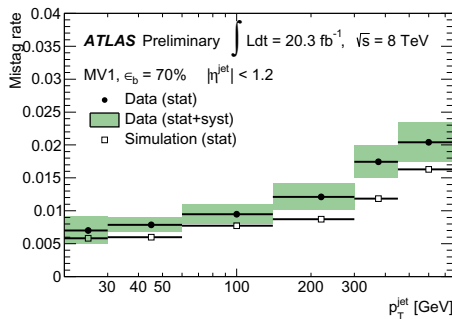


Figure 5: The mistag rate in data and simulation for the MV1 algorithm at 70% efficiency WP, obtained with the negative tag method, for jets with $|\eta| < 1.2$ [2].

7. Conclusion and outlook

The tagging efficiency of b -, c -, and light-flavor jets for the MV1 algorithm has been measured with LHC run 1 data using the 7 or 8 TeV dataset. The results are expressed in terms of data-to-simulation SFs, correcting the efficiencies in simulated events to those measured in data.

The best algorithms like MV1 provide for physics an unprecedented WP in which the b -jet tagging efficiency is at 70% while the average mistag rate is around 0.7%.

A combinatorial likelihood approach is applied to dileptonic $t\bar{t}$ events in order to measure the b -jet efficiency. The SFs have been measured and have a total uncertainty $\sim 2\%$ for $p_T^{\text{jet}} \sim 100$ GeV.

The W boson production in association with a c -quark is exploited to measure the tagging efficiency of c -jets. The total uncertainties ranging from 5% to 13%, increasing as the background rejection of the b -tagging algorithm increases.

Reconstructed D^* mesons associated with jets have been also used to measure the tagging efficiency of c -jets. The SFs are consistent with unity with uncertainties varying from 8% to 15% depending on p_T^{jet} .

The negative-tag method has been used to measure the light-jet efficiency (mistag rate) SFs. They are found to be larger than unity with a precision ranging between 15% and 43% depending on the p_T and η of jets.

As an outlook for Run 2, a new innermost layer of pixel detector called Insertable B-Layer (IBL) is now installed in the ATLAS detector. This will further improve the b -tagging performance because of improved tracking accuracy [5].

References

- [1] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.
- [2] ATLAS Collaboration, Calibration of the performance of b -tagging for c and light-flavour jets in the 2012 ATLAS data (ATLAS-CONF-2014-046), <http://cds.cern.ch/record/1741020>.
- [3] ATLAS Collaboration, Calibration of b -tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment (ATLAS-CONF-2014-004), <http://cds.cern.ch/record/1664335>.
- [4] ATLAS Collaboration, Calibration of the b -tagging efficiency for c jets with the ATLAS detector using events with a W boson produced in association with a single c quark (ATLAS-CONF-2013-109), <http://cds.cern.ch/record/1640162>.
- [5] ATLAS Collaboration, ATLAS Insertable B-Layer Technical Design Report (CERN-LHCC-2010-013. ATLAS-TDR-19), <http://cds.cern.ch/record/1291633>.